

ЗАДАЧІ ПОРОДЖЕННЯ ТА ФОРМАЛІЗАЦІЇ СТАЛИХ ТЕКСТІВ КАЗОК

Секало О.М., Шаронова Н.В.

*Національний технічний університет
"Харківський політехнічний інститут",
м. Харків, вул. Пушкінська, 79/2, тел. 707–64–60,
e-mail:sekalo.olga@gmail.com*

Проблема породження зв'язного тексту є дуже складною, оскільки текст є результатом взаємодії великої кількості факторів – лінгвістичних, психологічних, логічних тощо [1-4].

У процесі генерації тексту комп'ютером вчені виділяють три відносно незалежних етапи. На етапі *макропланування* комп'ютер приймає рішення, яка саме інформація з вхідних даних потрапить в текст і як вона буде організована. Після побудови плану тексту виконуються завдання етапу *мікропланування*. Його метою є складання плану окремих речень генерованого тексту з урахуванням загальної структури тексту. На етапі *мовного оформлення* ці плани перетворюються засобами лексики та граматики конкретної мови в граматичні структури, які потім трансформуються в речення тексту.

Необхідним є дотримання шляхів досягнення максимальної зрозумілості текстів, а саме: вибір вірного стилю; розробка семантичної структури тексту, перевірка тексту на наявність термінів та понять; введення в текст причинно-наслідкових зв'язків і речень.

Породження текстів казкової структури можна розділити на такі етапи:

1. Етап ідентифікації.

Суть етапу ідентифікації полягає у створенні системи породження чарівної казки, що спирається на деякий набір незмінних синтаксичних одиниць (інваріантів) і правил їх сполучуваності. Таким чином, тип бази знань для породження цього виду оповідних текстів та її основні складові визначаються тими інваріантними одиницями, які будуть виділені в процесі аналізу казок.

2. Етап концептуалізації.

При побудові системи породження оповідного тексту етап концептуалізації зводиться до виділення інваріантів казок і встановлення взаємозалежності між ними. Постійними та стійкими елементами казки є функції дійових осіб. Вони не залежать від того, ким і як виконуються. Виділяється близько тридцяти функцій, таких як: 1) відлучка; 2) заборона; 3) порушення заборони; 4) вивідання; 5) видача; 6) підступ; 7) пособництво; 8) нестача; 9) посередництво; 10) початок протидії та інші.

Іншими інваріантами чарівної казки є дійові особи. Виділено 10 дійових осіб: 1) Герой (ГР); 2) Антигерой (антагоніст, шкідник) (АГ); 3) Віщун (ВЩ); 4) Дарувальник (ДР); 5) Помічник (ПМ); 6) Антипомічник (АП); 7) Дурень (ДР); 8) Антидарувальник (АД); 9) Нагорода (НД); 10) Перешкода (ПР).

Кожна дійова особа в казці може бути описана трьома характеристиками: атрибутами, вчинками та зустрічами.

Атрибути – це сукупність всіх зовнішніх якостей персонажів: їх вік, стать, стан, зовнішній вигляд, особливості цього вигляду та інші індивідуальні риси.

Вчинки – це дії, що здійснюються дійовими особами казки. Наприклад, «*піти*», «*відправитися*», «*битися*», «*дістати*» тощо.

Зустрічі дійових осіб безпосередньо пов'язані з їхніми вчинками. Наприклад, дарувальник дасть Герою Помічника (*щуку, коня, вовка*), або Дарувальник дасть Герою пораду про те, як діяти далі.

База знань всієї системи казки складається з трьох самостійних баз знань: експозиції, тіла казки і постпозиції [4].

Експозиція, або початкова частина казки, являє собою деяку кінцеву множину частин фраз, цілих фраз або послідовностей фраз. Такі фрази можуть виражати стандартні формули початку казки. Наприклад: *У деякому царстві, у деякій державі.*

Тіло казки в найзагальнішому вигляді є кінцева послідовність зустрічей дійових осіб, пов'язаних сполучними фразами. Наприклад: *Іде він і бачить ...; Сів він на килим-літак і полетів ...; І пішов він далі...*

Постпозиції – це завершальна частина казки. Є певні традиційні формули кінця казки. Наприклад: *Я там був, мед, пиво пив, по вусах текло, а в рот не попало; Цар зрадив і видав за нього свою дочку.*

3. Етап формалізації.

На етапі формалізації складається наступний, більш загальний вид тіла створюваної казки. Зміст казки може бути представлений у вигляді послідовності коротких фраз. Породження тіла казки починається з вибору дійових осіб. Далі, вказуються дієслівні фрейми, а саме дії та вчинки персонажів.

Тестування системи породження казки полягає в перевірці великого числа породжуваних формул казок з використанням різних дійових осіб.

Наведена нижче казка була синтезована електронно-обчислювальною машиною в ході виконання створеної програми TALE (мова програмування LISP).

Жив-був цар, мав дочку. Налетів змій, забрав доньку. Послав цар Іванка відшукати доньку, обіцяв доньку в дружини. Відправився Іванко шукати доньку. Йде Іванко, бачить печеру, увійшов Іванко в печеру, дивиться Іванко, сидить донька. Сховався Іванко, прилетів змій, схопив Іванко змія, покликав мандрівника, вбили змія. Забрав Іванко доньку, повернувся назад. Отримав Іванко доньку в дружини.

Даний фрагмент тексту був сформований з урахуванням запропонованого алгоритму і цілком відповідає сталому сценарію породження казки [4].

При детальному аналізі наведеного фрагменту тексту казки необхідно зазначити загальний недолік процедури формалізації казки. Із наведеного вище прикладу можна помітити, що текст, сформований із використанням алгоритму, втрачає художнє забарвлення та деталізацію подій, що є недоліком при обробці художнього тесту. Тому дана концепція формалізації текстів казок потребує подальшого удосконалення. Між тим, створений текст казки є гарним початком



для подальшого опрацювання тексту не машинними засобами, а людиною, яка спеціалізується на написанні художніх творів.

Розгляд даної задачі є початком для подальших детальних досліджень запропонованої теми. Існує думка, що формалізація художнього тексту є зовсім нездійсненною. Адже, найбільш складними для розуміння є тексти, що описують взаємини і вчинки активних дійових осіб, оскільки в цьому випадку мають створюватися та повинні зберігатися великі бази знань про світ. Такі системи повинні мати досконалі механізми виведення з наявних знань, що визначають окремих суб'єктів, їх властивості, дії і вчинки, а також мотиви цих вчинків. На даному етапі розвитку комп'ютерних та лінгвістичних наук ця задача не є вирішеною. І хоча зазначена тема на цей час не має значного практичного застосування, з урахуванням швидкого прогресу у формалізації природної мови можна припустити, що дослідникам також вдасться розробити систему породження художнього тексту [2, 3].

Список літератури

1. Бремон К. Структурное изучение повествовательных текстов после В. Проппа / К. Бремон // Семиотика. – М. : Радуга, 1983. – 436 с.
2. Мелетинский Е.М. Структурно-типологическое изучение сказки / Е.М. Мелетинский // Пропп В.Л. Морфология сказки. – М. : Наука, 1969. – 166 с.
3. Новик Е.С. Система персонажей русской волшебной сказки / Е.С. Новик // Типологические исследования по фольклору. – М. : Наука, 1975. – 247 с.
4. Zubov A.V. Основы искусственного интеллекта для лингвистов: Учеб. пособие / А.В. Zubov, И.И. Zubova. – М. : Университетская книга, Логос, 2007. – 320 с.